

CURVATURE TERMS IN SMALL TIME HEAT KERNEL EXPANSION FOR A MODEL CLASS OF HYPOELLIPTIC HÖRMANDER OPERATORS

DAVIDE BARILARI[†] AND ELISA PAOLI[‡]

ABSTRACT. We consider the heat equation associated with a class of second order hypoelliptic Hörmander operators with constant second order term and linear drift. We describe the possible small time heat kernel expansion on the diagonal giving a geometric characterization of the coefficients in terms of the divergence of the drift field and the curvature-like invariants of the optimal control problem associated with the diffusion operator.

CONTENTS

1. Introduction	1
2. Linear quadratic optimal control problems	6
3. The flag and growth vector of an admissible curve	7
4. Geodesic cost and curvature invariants	8
5. Small time asymptotics at an equilibrium point	9
6. Small time asymptotics out of the equilibrium	12
Appendix A. On the fundamental solution	15
References	17

1. INTRODUCTION

The interaction between analysis and geometry in the study of the heat equation in several smooth (and more recently, non-smooth) contexts has attracted an increasing attention during the last decades.

On a Riemannian manifold, many results have been obtained relating the small time asymptotics of the fundamental solution of the heat equation, the so-called *heat kernel*, to geometric quantities such as distance [23], cut and conjugate locus [17, 18], curvature invariants [11]. The extension of these results to non-Riemannian situations (from the geometric viewpoint) or non-elliptic operators (from the viewpoint of PDE), when possible, is non trivial: some results have been obtained relating the hypoelliptic heat kernel with its associated Carnot-Carathéodory distance [15, 16] and its cut locus [5, 6], but much less is known concerning the relation with curvature. This is also related to the difficulty of defining a general notion of curvature associated to Carnot-Carathéodory metrics.

One of the most celebrated results in the Riemannian setting, which is by now classical and in the spirit of the investigation presented in this paper, reads as follows. Denote by $p(t, x, y)$ the heat kernel associated with the Laplace-Beltrami

[†]INSTITUT DE MATHÉMATIQUES DE JUSSIEU-PARIS RIVE GAUCHE UMR CNRS 7586, UNIVERSITÉ PARIS-DIDEROT, BATIMENT SOPHIE GERMAIN, CASE 7012, 75205 PARIS CEDEX 13, FRANCE

[‡]SISSA, VIA BONOMEA 265, TRIESTE, ITALY

E-mail addresses: davide.barilari@imj-prg.fr, epaoli@sissa.it.

Date: October 19, 2015.

operator Δ_g on a Riemannian manifold (M, g) , then the coefficients appearing in the small time heat kernel expansion on the diagonal

$$p(t, x_0, x_0) = \frac{1}{t^{n/2}} \left(\sum_{i=0}^m a_i(x_0) t^i + O(t^m) \right)$$

contain information about the curvature of the manifold at the point x_0 , i.e., all $a_i(x_0)$ can be written as universal polynomials of the Riemann tensor and its covariant derivatives computed at the point x_0 (see for instance [11, 20]).

More generally, if one is interested in the study of the heat kernel associated with a smooth second order elliptic operator L , possibly with drift, one can extract the good Riemannian metric g from the principal symbol of the operator L , such that the on-the-diagonal heat kernel expansion contains geometric information about curvature associated to the metric g and the local structure of the drift at the point. For instance, in [12], it is proved that for the operator $L = \Delta_g + X_0$ associated to a Riemannian manifold (M, g) the corresponding heat kernel satisfies for $t \rightarrow 0$

$$(1) \quad p(t, x_0, x_0) = \frac{1}{(4\pi t)^{n/2}} \left(1 - \left(\frac{\operatorname{div}(X_0)}{2} + \frac{\|X_0(x_0)\|^2}{2} - \frac{S(x_0)}{6} \right) t + O(t^2) \right),$$

where S is the scalar curvature of the Riemannian metric g .

Let us now consider the second order Hörmander-type operator

$$(2) \quad L = \frac{1}{2} \sum_{i=1}^k X_i + X_0$$

satisfying the *Hörmander hypoellipticity condition*

$$(wHC) \quad \operatorname{Lie}\{(\operatorname{ad} X_0)^j X_i \mid i = 1, \dots, k, j \in \mathbb{N}\} \Big|_x = T_x M, \quad \forall x \in M.$$

(Here $(\operatorname{ad} X)Y = [X, Y]$ and $\operatorname{Lie} \mathcal{F}$ denotes the smallest Lie algebra containing a family of vector fields \mathcal{F}).

Condition (wHC) guarantees the existence of a smooth heat kernel $p(t, x, y)$ for the heat equation associated with L defined in (2). Condition (wHC) is also sometimes referred as *weak Hörmander condition*, in contrast to its *strong* version (sHC) where only the vector field appearing in the second order term of L are required to generate the tangent space

$$(sHC) \quad \operatorname{Lie}\{X_i \mid i = 1, \dots, k\} \Big|_x = T_x M, \quad \forall x \in M.$$

As soon as one removes the ellipticity assumptions on L , the geometric content of the coefficients in the asymptotic expansion of the heat kernel (and indeed even the structure of this asymptotics) is much less understood.

For instance, when the operator L is elliptic, the decay on the diagonal of $p(t, x, x)$ for $t \rightarrow 0$ is always of polynomial type, for every drift X_0 (cf. (1)). On the other hand, when L is only hypoelliptic, the drift plays a crucial role in the structure of the asymptotics.

Due to hypoellipticity, the diffusion generated by the second order term acts in all directions but in an anisotropic way. If the drift field is a smooth section of $\mathcal{G}_2(x) := \operatorname{span}_x\{X_i, [X_j, X_k] : i, j, k \geq 1\}$, then it modifies the coefficients of the diffusion, but only at higher terms and does not affect the order of the asymptotics. On the other hand, if the drift field points outside \mathcal{G}_2 , it has a stronger effect on the diffusion and its flow could not be compensated by the fields X_1, \dots, X_k , generating a rapid decrease of the heat at the point (see also [9]). This different behavior, given by second order Lie brackets, can be heuristically explained by the fact that the second order part of the operator reproduces a Brownian motion, which moves as

\sqrt{t} , while the drift field has velocity t , then if X_0 points outside \mathcal{G}_2 the diffusion generated by the second order part is too slow.

More precisely, when the drift vector field is either zero or horizontal, i.e., $X_0 = \sum f_i X_i$ for some smooth functions f_i , conditions (wHC) and (sHC) coincide and Ben Arous proved in [3] that the decay on the diagonal is polynomial¹

$$p(t, x_0, x_0) \sim t^{-Q/2}$$

where Q is the Hausdorff dimension associated to the Carnot-Carathéodory metric defined by X_1, \dots, X_k . More generally, in presence of a non-zero drift and under the weak Hörmander condition assumption (wHC), in [19] it is proved that either

$$p(t, x_0, x_0) \sim t^{-\mathcal{N}/2}$$

or $p(t, x_0, x_0)$ has a faster decay, even exponential for every x_0 that is an equilibrium point for the drift. Here the difference in the behaviour is characterized by the controllability of an associated control problem. Moreover \mathcal{N} is an integer that depends on the structure of the Lie algebra (wHC), that coincides with Q when X_0 is horizontal.

When $X_0(x_0) \neq 0$ does not belong to $\mathcal{G}_2(x_0)$, then the flow of the drift locally interacts with the heat diffusion given by the second order part and generates an exponential decay, see [9, 10].

Concerning the geometric meaning of the coefficients of the asymptotic expansion, few results are available and only when the drift field is either zero or horizontal. In [4] it is computed the first term of the asymptotic for 3D contact structures, where an invariant κ of the sub-Riemannian structure playing the role of the curvature appears. Concerning higher dimensional structures, to our best knowledge, the only known results are [21] for the case of a Sasakian manifold (where the trace of the Tanaka Webster curvature appears) and the case of the two higher dimensional model spaces: CR spheres [7] and quaternionic Hopf fibrations [8].

In this paper we perform the first step in the understanding the structure of the asymptotic expansion at the diagonal in the case of an operator with general drift. In particular we are interested in the geometric characterization of its coefficients, for the model class of Hörmander operators in \mathbb{R}^n where X_0 is a linear field and X_i are constant and linearly independent:

$$X_0 = \sum_{j,h=1}^n a_{jh} x_h \partial_{x_j}, \quad X_i = \sum_{j=1}^n b_{ij} \partial_{x_j}, \quad i = 1, \dots, k.$$

If we denote by $A = (a_{jh})$ and $B = (b_{ij})$ respectively the $n \times n$ and $n \times k$ matrices of the coefficients the operator L assumes the form

$$(3) \quad L = \sum_{j=1}^n (Ax)_j \frac{\partial}{\partial x_j} + \frac{1}{2} \sum_{j,h=1}^n (BB^*)_{jh} \frac{\partial^2}{\partial x_j \partial x_h}.$$

The (weak) Hörmander condition for the operator (3) is equivalent to the assumption that there exists a minimal $m \in \mathbb{N}$ such that

$$(4) \quad \text{rk}[B, AB, A^2B, \dots, A^{m-1}B] = n.$$

The heat equation associated to L admits a smooth fundamental solution $p(t, x, y) \in C^\infty(\mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n)$ that can be computed explicitly as follows

$$(5) \quad p(t, x, y) = \frac{e^{-\frac{1}{2}(y - e^{tA}x)^* D_t^{-1} (y - e^{tA}x)}}{(2\pi)^{n/2} \sqrt{\det D_t}}$$

¹Here and in what follows $f(t) \sim g(t)$ means $f(t) = g(t)(C + o(1))$ for some $C \neq 0$ and $t \rightarrow 0$.

where

$$D_t = e^{tA} \left(\int_0^t e^{-\tau A} B B^* e^{-\tau A^*} d\tau \right) e^{tA^*}.$$

By condition (4), the matrix D_t is invertible for every $t > 0$.

These operators are the simplest class of hypoelliptic, but not elliptic, operators satisfying (wHC) and are classical in the literature, starting from the pioneering work of Hörmander [13] (see also [14] for a detailed discussion on this class of operators). Their fundamental solution is the probability density of the stochastic process ξ_t solution of

$$d\xi_t = A\xi_t dt + Bdw(t),$$

where $w(t)$ is a k -dimensional Brownian motion. For the reader's convenience we present in Appendix the construction of the fundamental solution and its relation to the associated control problem described below.

Indeed, as already pointed out by Stroock and Varadhan [22] in the study of the support of the diffusion, the properties of ξ_t are strongly related with the solutions to the control problem

$$(6) \quad \dot{x}(t) = Ax(t) + Bu(t), \quad x \in \mathbb{R}^n, u \in \mathbb{R}^k.$$

The relation between the solution of the heat equation and the associated control problem is classical, see for instance [12]. In this paper we further investigate this relation and we show how geometric invariants associated to the *cost functional*

$$J_T(u) = \frac{1}{2} \int_0^T |u(s)|^2 ds$$

reveals some geometric-like properties of the heat kernel. In other words, for every fixed $x_1, x_2 \in \mathbb{R}^n$ and $T > 0$, one is interested in computing

$$S_T(x_1, x_2) := \inf \{ J_T(u) : u \in L^\infty([0, T]; \mathbb{R}^k), x_u(0) = x_1, x_u(T) = x_2 \}.$$

where $x_u(\cdot)$ is the solution of (6) associated with the control u . The condition (4) (also known as *Kalman condition*) ensures that the control system (6) is controllable, i.e., $S_T(x_1, x_2) < +\infty$ for all $x_1, x_2 \in \mathbb{R}^n$ and $T > 0$.

For $x_0 \in \mathbb{R}^n$ fixed, let $x_{\bar{u}}(t)$ be an optimal trajectory starting at x_0 , i.e., a minimizer of the cost functional. The *geodesic cost* associated with $x_{\bar{u}}$ is the family of functions

$$c_t(x) := -S_t(x, x_{\bar{u}}(t)) \quad \text{for } t > 0, x \in \mathbb{R}^n.$$

From the asymptotics of c_t , one can highlight some “curvature-like” invariants of the cost, which define a family of symmetric operators

$$\mathcal{I} : \mathbb{R}^k \rightarrow \mathbb{R}^k, \quad \mathcal{Q}^{(i)} : \mathbb{R}^k \rightarrow \mathbb{R}^k, \quad i \geq 0.$$

These operators, that are in principle associated with an optimal trajectory, in the case of a linear-quadratic optimal control problem are constant.

The operator \mathcal{I} is connected to the flag generated by the brackets along the optimal trajectory. The operators $\mathcal{Q}^{(i)}$ play the role of curvature invariants for the optimal control problem (see Section 4 and [1] for more details).

We recall here the geometric meaning of the trace of \mathcal{I} that is crucial in what follows. Define the filtration $E_1 \subset E_2 \subset \dots \subset E_m = \mathbb{R}^n$ as

$$(8) \quad E_i = \text{span}\{A^j Bx \mid x \in \mathbb{R}^k, 0 \leq j \leq i-1\}.$$

Then

$$(9) \quad \mathcal{N} := \text{tr}(\mathcal{I}) = \sum_{i=0}^{m-1} (2i-1)(\dim E_{i+1} - \dim E_i).$$

1.1. Main results. When x_0 is an equilibrium of the drift field, we can compute and characterize all the coefficients in the small time asymptotic expansion, providing a characterization of the coefficients that is analogous to the one obtained on a Riemannian manifold.

Theorem 1.1. *Assume that $Ax_0 = 0$. Then*

$$p(t, x_0, x_0) = \frac{t^{-\mathcal{N}/2}}{(2\pi)^{n/2} \sqrt{c_0}} \left(\sum_{i=0}^m a_i t^i + O(t^m) \right), \quad \text{for } t \rightarrow 0,$$

where $\mathcal{N} = \text{tr}(\mathcal{I})$ is defined in (9) and c_0 is a positive constant. Moreover there exists universal polynomials P_i of degree i such that

$$a_i = P_i(\text{tr} A, \text{tr} \mathcal{Q}^{(0)}, \dots, \text{tr} \mathcal{Q}^{(i-2)}).$$

In particular for $i = 1, 2, 3$, we have

$$\begin{aligned} a_1 &= -\frac{\text{tr} A}{2}, & a_2 &= \frac{(\text{tr} A)^2}{8} + \frac{\text{tr} \mathcal{Q}^{(0)}}{4}, \\ a_3 &= -\frac{\text{tr} \mathcal{Q}^{(1)}}{12} - \frac{\text{tr} A \text{tr} \mathcal{Q}^{(0)}}{8} - \frac{(\text{tr} A)^3}{48}. \end{aligned}$$

We stress that the explicit structure of any higher order coefficient can be a priori computed by a simple Taylor expansion, as it follows from the proof, cf. Section 5.

More in general, one has an expansion of $p(t, x, y)$, at every pair of points x, y , relating the heat kernel with the optimal cost functional and the same geometric coefficients of Theorem 1.1.

Corollary 1.2. *For any pair of points, $x, y \in \mathbb{R}^n$,*

$$p(t, x, y) = \frac{t^{-\mathcal{N}/2}}{(2\pi)^{n/2} \sqrt{c_0}} e^{-S_t(x, y)} \left(\sum_{i=0}^m a_i t^i + O(t^m) \right), \quad \text{for } t \rightarrow 0,$$

where the coefficients a_i are characterized as in Theorem 1.1.

This corollary is a direct consequence of the previous theorem and is proved in Section 5.2.

Next we consider the case when x_0 is not a zero of the drift field. In this case, as explained, one can observe different behaviors depending on the smallest level of the filtration (8) to which the vector Ax_0 belongs. Indeed the optimal cost to remain fixed in the point x_0 is strictly positive and the asymptotics depends on the exponential in Theorem 1.2.

Theorem 1.3. *Assume that $Ax_0 \neq 0$. Then*

(i) *If $Ax_0 \in E_1$, then we have the polynomial decay*

$$p(t, x_0, x_0) = \frac{t^{-\mathcal{N}/2}}{(2\pi)^{n/2} \sqrt{c_0}} \left[1 - \left(\frac{\text{tr} A}{2} + \frac{|Ax_0|^2}{2} \right) t + O(t^2) \right], \quad \text{for } t \rightarrow 0.$$

(ii) *If $Ax_0 \in E_i \setminus E_{i-1}$ for some $i > 1$, then $p(t, x_0, x_0)$ has exponential decay to zero. More precisely there exists $C > 0$ such that*

$$p(t, x_0, x_0) = \frac{t^{-\mathcal{N}/2}}{(2\pi)^{n/2} \sqrt{c_0}} \exp \left(-\frac{C + O(t)}{t^{2i-3}} \right), \quad \text{for } t \rightarrow 0.$$

We stress that claim (i) is the analogue of the expansion (1): here no scalar curvature appears since all brackets of horizontal fields are zero. For the same reason, in this case the space \mathcal{G}_2 coincides with E_1 and in claim (ii), when $Ax_0 \notin E_1$, we indeed observe an exponential decay. The proof of Theorem 1.3 is presented in Section 6.

2. LINEAR QUADRATIC OPTIMAL CONTROL PROBLEMS

Let us consider the optimal control problem associated with the operator L

$$(10) \quad \begin{cases} \dot{x} = Ax + Bu \\ J_T(u) = \frac{1}{2} \int_0^T \sum_{i=1}^k |u_i(s)|^2 ds \rightarrow \min \end{cases}$$

Here $u \in L^\infty([0, T]; \mathbb{R}^k)$ is the control and J_T is the optimal cost to be minimized. A curve $x(t)$ in \mathbb{R}^n is called *admissible* for the control problem (10) if there exists a control function $u \in L^\infty([0, T]; \mathbb{R}^k)$ such that $\dot{x}(t) = Ax(t) + Bu(t)$ for a.e. $t \in [0, T]$.

The solution of the differential equation (10) corresponding to the control u will be denoted by $x_u : [0, T] \rightarrow \mathbb{R}^n$ and for a fixed initial point $x_1 \in \mathbb{R}^n$ is given by:

$$(11) \quad x_u(t) = e^{tA}x_1 + e^{tA} \int_0^t e^{-\tau A} Bu(\tau) d\tau.$$

Among all trajectories x_u starting at x_1 and arriving in a point $x_2 \in \mathbb{R}^n$ in time T we want to minimize the cost functional J_T : for every fixed $x_1, x_2 \in \mathbb{R}^n$ and $T > 0$, we define the *value function*

$$(12) \quad S_T(x_1, x_2) := \inf\{J_T(u) | u \in L^\infty([0, T]; \mathbb{R}^k), x_u(0) = x_1, x_u(T) = x_2\}.$$

A control \bar{u} that realizes the minimum in (12) is called an *optimal control*, and the corresponding trajectory $x_{\bar{u}} : [0, T] \rightarrow \mathbb{R}^n$ is called an *optimal trajectory* of the control problem (10).

It is well-known (see for example [2]) that the optimal trajectories of the control problem (10) can be obtained as the projection of the solutions of an Hamiltonian system in $T^*\mathbb{R}^n$. Namely, let

$$H(p, x) = p^*Ax + \frac{1}{2}p^*BB^*p, \quad \forall (p, x) \in T^*\mathbb{R}^n,$$

be the Hamiltonian function associated with the optimal control problem. All the optimal trajectories are the projection $x(t)$ of the solution $(p(t), x(t)) \in T^*\mathbb{R}^n \cong \mathbb{R}^{2n}$ of the Hamiltonian system associated with H

$$\begin{cases} \dot{p} = -A^*p \\ \dot{x} = Ax + BB^*p. \end{cases}$$

Moreover, the control realizing the optimal trajectory is uniquely recovered by $\bar{u}(t) = B^*p(t)$. Thus the solution corresponding to the initial condition $(p_0, x_0) \in T_{x_0}^*\mathbb{R}^n$ can be found explicitly

$$(13) \quad \begin{cases} p(t) = e^{-tA^*}p_0 \\ x(t) = e^{tA} \left(x_0 + \int_0^t e^{-\tau A} BB^* e^{-\tau A^*} d\tau p_0 \right). \end{cases}$$

Let us denote by Γ_t the matrix

$$(14) \quad \Gamma_t := \int_0^t e^{-\tau A} BB^* e^{-\tau A^*} d\tau.$$

By Kalman's condition (4), it follows that Γ_t is invertible for every $t > 0$.

Remark 2.1. Fix $x_1, x_2 \in \mathbb{R}^n$ and $T > 0$. By the explicit formulas (13) there exists a unique initial covector p_0 such that the corresponding extremal $x(t)$ satisfies $x(0) = x_1$ and $x(T) = x_2$. It is equal to

$$p_0 = \Gamma_T^{-1} (e^{-TA}x_2 - x_1).$$

Since the optimal control is given by $\bar{u}(t) = B^*p(t)$, we can also write the optimal cost to go from x_1 to x_2 , namely

$$S_T(x_1, x_2) = \frac{1}{2}p_0^* \Gamma_T p_0.$$

It follows that the cost function is smooth in $(T, x_1, x_2) \in \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n$, where $\mathbb{R}^+ = \{T \in \mathbb{R} : T > 0\}$.

3. THE FLAG AND GROWTH VECTOR OF AN ADMISSIBLE CURVE

Let $x_u : [0, T] \rightarrow \mathbb{R}^n$ be an admissible curve such that $x_u(0) = x_0$, associated with the control u . Let $P_{0,t}$ be the flow defined by u , i.e., for every $y \in \mathbb{R}^n$

$$P_{0,t}(y) := x_u(t; y) \quad \text{s.t.} \quad x_u(0; y) = y.$$

At any point of \mathbb{R}^n we split the tangent space $T_x \mathbb{R}^n \cong \mathbb{R}^n = D \oplus D^\perp$, where D is the k -dimensional subspace generated by the columns of B and D^\perp is its orthogonal complement, and we define the following family of subspaces of $T_{x_0} \mathbb{R}^n$:

$$\mathcal{F}_{x_u}(t) := (P_{0,t})_*^{-1} D \subset T_{x_0} \mathbb{R}^n.$$

In other words, the family $\mathcal{F}_{x_u}(t)$ is obtained by translating in $T_{x_0} \mathbb{R}^n$ the subspace D along the trajectory x_u using the flow $P_{0,t}$.

Definition 3.1. The *flag of the admissible curve* $x_u(t)$ is the sequence of subspaces

$$\mathcal{F}_{x_u}^i(t) := \text{span} \left\{ \frac{d^j}{dt^j} v(t) \mid v(t) \in \mathcal{F}_{x_u}(t) \text{ smooth}, j \leq i-1 \right\} \subset T_{x_0} \mathbb{R}^n, \quad i \geq 1.$$

By construction, this is a filtration of $T_{x_0} \mathbb{R}^n$, i.e., $\mathcal{F}_{x_u}^i(t) \subset \mathcal{F}_{x_u}^{i+1}(t)$, for all $i \geq 1$.

Definition 3.2. Denote by $k_i(t) := \dim \mathcal{F}_{x_u}^i(t)$. The *growth vector of the admissible curve* $x_u(t)$ is the sequence of integers

$$\mathcal{G}_{x_u}(t) = \{k_1(t), k_2(t), \dots\}.$$

Remark 3.3. For any $s \in [0, T]$ we can define the family of subspace $\mathcal{F}_{x_{u,s}}(t)$ associated with the admissible curve x_u starting at time s . Namely, let $P_{s,t}$ be the flow defined by u starting at time $s < t$, i.e., for every $y \in \mathbb{R}^n$

$$P_{s,t}(y) := x_u(t; y) \quad \text{s.t.} \quad x_u(s; y) = y,$$

and let $x_{u,s}(t) := x_u(s+t)$ be the shifted curve by time s . Then we introduce the family of subspaces $\mathcal{F}_{x_{u,s}}(t) := (P_{s,s+t})_*^{-1} \mathbb{R}^k$ with base point $x_u(s)$.

The relation $\mathcal{F}_{x_{u,s}}(t) = (P_{0,s})_* \mathcal{F}_{x_u}(s+t)$ implies that the growth vector of the original curve at time t can be equivalently computed via the growth vector at time 0 of the shifted curve $x_{u,t}$, i.e., $k_i(t) = \dim \mathcal{F}_{x_{u,t}}^i(0)$ and $\mathcal{G}_{x_u}(t) = \mathcal{G}_{x_{u,t}}(0)$.

The growth vector of a curve x_u at time 0 can be easily computed. Indeed, by the explicit expression (11) of the flow of u , the flag of the curve x_u starting from a point x_0 is

$$\mathcal{F}_{x_u}^i(0) = \text{span}\{B, AB, \dots, A^{i-1}B\}.$$

In particular the flag at time 0 is independent on the control u and the initial point x_0 . By Remark 3.3 the growth vector of any curve x_u is then independent also from the time and is equal to $\mathcal{G}_{x_u}(t) = \{k_1, k_2, \dots\}$, where the indices k_i are

$$k_i := \dim \text{span}\{B, AB, \dots, A^{i-1}B\}.$$

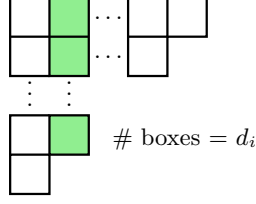


FIGURE 1. Young diagram

By Kalman's condition (4) there exists a minimal integer $1 \leq m \leq n$ such that $k_m = n$. We call such m the *step* of the control problem (independent of the admissible curves). Moreover notice that $k_1 = k$.

Lemma 3.4. *Let $d_i := k_i - k_{i-1}$. Then $d_1 \geq d_2 \geq \dots \geq d_m$.*

Proof. The linear map $\hat{A} : \mathcal{F}_{x_u}^i(t) \rightarrow \mathcal{F}_{x_u}^{i+1}(t)/\mathcal{F}_{x_u}^i(t)$ defined by $\hat{A}v := Av$ for every $v \in \mathcal{F}_{x_u}^i(t)$ is surjective and $\text{Ker}\hat{A} = \mathcal{F}_{x_u}^{i-1}(t)$. Then

$$\dim \mathcal{F}_{x_u}^i(t) - \dim \mathcal{F}_{x_u}^{i-1}(t) \geq \dim \mathcal{F}_{x_u}^{i+1}(t) - \dim \mathcal{F}_{x_u}^i(t),$$

which concludes the proof. \square

To any curve x_u we associate a tableau with m columns of length d_i for $i = 1, \dots, m$. By the previous Lemma, the height of the columns is decreasing from left to right. We call n_j the length of the j -th row, for $j = 1, \dots, k$ (for example $n_1 = m$, see Figure 1).

4. GEODESIC COST AND CURVATURE INVARIANTS

Definition 4.1. Let $x_0 \in \mathbb{R}^n$ and fix an optimal trajectory $x_{\bar{u}}(t)$ starting at x_0 of the optimal control problem (10). The *geodesic cost* associated with $x_{\bar{u}}$ is the family of functions $\{c_t\}_{t>0}$ defined by

$$c_t(x) := -S_t(x, x_{\bar{u}}(t)), \quad x \in \mathbb{R}^n,$$

where S_t is the optimal cost function defined in (12).

Notice that thanks to Remark 2.1 and the smoothness of optimal trajectories, the geodesic cost is smooth in $\mathbb{R}^+ \times \mathbb{R}^n$.

Moreover, for any $t > 0$ and $x \in \mathbb{R}^n$ there exists a unique minimizer of the cost functional among all the trajectories that connect x with $x_{\bar{u}}(t)$ in time t . By the explicit expressions of the extremals in (13) and of the optimal control \bar{u} , we can write the explicit formula:

$$c_t(x) = -\frac{1}{2}p_0^* \Gamma_t p_0 + p_0^*(x - x_0) - \frac{1}{2}(x - x_0)^* \Gamma_t^{-1}(x - x_0),$$

where $x_{\bar{u}}$ is a solution of the Hamiltonian system with initial data (p_0, x_0) and Γ_t is the invertible matrix defined in (14).

Then we define the following family of quadratic forms, $\mathcal{Q}(t)$, on \mathbb{R}^k :

$$(15) \quad \mathcal{Q}(t) := B^* \left(d_{x_0}^2 \dot{c}_t \right) B = -\frac{d}{dt} B^* \Gamma_t^{-1} B.$$

This family of operators is the linear quadratic counterpart of the more general family of operators introduced in [1, Chapter 4] for the wider class of non linear control systems that are affine in the control.

Remark 4.2. Actually the family of operators $\mathcal{Q}(t)$ does not depend on the initial data (p_0, x_0) of the optimal trajectory and is the same for any geodesic. This is saying that it is an intrinsic object of the pair control system and cost.

Moreover, let us stress that $\mathcal{Q}(t)$ is an intrinsic object of the optimal control problem, i.e., it does not depend on the chosen coordinate on \mathbb{R}^n .

Indeed let $y = Cx$, with C an $n \times n$ invertible matrix, be a linear change of coordinates on \mathbb{R}^n . In the new coordinates the dynamical system (10) is rewritten as

$$\dot{y} = \tilde{A}y + \tilde{B}u,$$

where $\tilde{A} = CAC^{-1}$ and $\tilde{B} = CB$. Since the matrix C is invertible the dimensions d_i remain unchanged, and in particular $\text{rk } B = \text{rk } \tilde{B} = k$. The matrix $\tilde{\Gamma}_t$ for the new coordinate system can be easily recovered by definition (14) and we find $\tilde{\Gamma}_t = C\Gamma_t C^*$. This implies in particular that the definition of $\mathcal{Q}(t)$ in (15) is independent on the coordinates. See also [1] for a general discussion about the geodesic cost.

The next theorem shows the asymptotic behavior of the family $\mathcal{Q}(t)$ for small $t > 0$. The proof of this result, in its general setting, can be found in [1, Chapter 4].

Theorem 4.3 (Theorem A in [1]). *Let $x_{\bar{u}} : [0, T] \rightarrow \mathbb{R}^n$ be an optimal trajectory. The function $t \mapsto t^2 \mathcal{Q}(t)$ defined in Eq. (15) can be extended to a smooth family of symmetric operators on \mathbb{R}^n for all $t \geq 0$. In particular, for every fixed $h \in \mathbb{N}$, one has the following Laurent expansion*

$$(16) \quad \mathcal{Q}(t) = \frac{1}{t^2} \mathcal{I} + \sum_{i=0}^h \mathcal{Q}^{(i)} t^i + O(t^{h+1}), \quad t > 0.$$

Moreover all the matrices \mathcal{I} and $\mathcal{Q}^{(i)}$ are symmetric and

$$\text{tr } \mathcal{I} = \sum_{i=1}^k n_i^2 = \sum_{j=1}^m (2j-1) d_j.$$

The expansion (16) defines a sequence of symmetric operators (or matrices) \mathcal{I} and $\mathcal{Q}^{(i)}$, for $i \in \mathbb{N}$. These operators are canonically associated to the optimal control problem.

5. SMALL TIME ASYMPTOTICS AT AN EQUILIBRIUM POINT

In this section we prove Theorem 1.1, concerning the small time asymptotics of the fundamental solution $p(t, x, x)$ at a point $x \in \text{Ker } A$. We will write this expansion in terms only of the drift field and the invariant $\mathcal{Q}^{(i)}$ introduced in Theorem 4.3.

Remark 5.1. The order of the asymptotics, \mathcal{N} , was already found in [19] for a wider class of hypoelliptic operators. In particular, for the linear case the number \mathcal{N} is determined only by the numbers $k_i = \text{rk}\{B, AB, \dots, A^{i-1}B\}$, for every $1 \leq i \leq m-1$, and $k_0 = 0$. Indeed \mathcal{N} can be rewritten as follows

$$\mathcal{N} := \sum_{i=1}^m (2i-1)(k_i - k_{i-1}).$$

This number is actually equal to the trace of \mathcal{I} , given in Theorem 4.3, since by the classical identity $\sum_{i=1}^n (2i-1) = n^2$ we have

$$\text{tr } \mathcal{I} = \sum_{i=1}^k n_i^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (2j-1) = \sum_{j=1}^m (2j-1)(k_j - k_{j-1}) = \mathcal{N}.$$

Recall now the expression of the fundamental solution $p(t, x, y)$ given in (5):

$$p(t, x, y) = \frac{e^{-\frac{1}{2}(y-e^{tA}x)^* D_t^{-1}(y-e^{tA}x)}}{(2\pi)^{n/2} \sqrt{\det D_t}},$$

where the matrix D_t is defined as

$$(17) \quad D_t = e^{tA} \int_0^t e^{-\tau A} B B^* e^{-\tau A^*} d\tau e^{tA^*} = e^{tA} \Gamma_t e^{tA^*}.$$

A change of variable in the integral defining Γ_t gives easily that $D_t = -\Gamma_{-t}$.

If a point x is an equilibrium point for the drift field (i.e., $x \in \text{Ker} A$), then the asymptotics of $p(t, x, x)$ on the diagonal is determined uniquely by the Taylor expansion of $\sqrt{\det D_t}^{-1}$, since $e^{tA}x = x$ for every t .

Let $d(t) := \det D_t$, then $d(t)$ satisfies

$$(18) \quad d'(t) = d(t) \text{tr}(D_t^{-1} \dot{D}_t) = d(t)(2\text{tr}(A) + \text{tr}(B^* D_t^{-1} B)).$$

Moreover, since $d(t)$ has a simple pole at $t = 0$ of order \mathcal{N} we can write the determinant as $d(t) = t^{\mathcal{N}} f(t)$, for some smooth function f non-vanishing at zero. Substituting this expression in (18) one gets

$$(19) \quad \frac{d'(t)}{d(t)} = \frac{\mathcal{N}}{t} + \frac{f'(t)}{f(t)} = 2\text{tr}(A) + \text{tr}(B^* D_t^{-1} B).$$

Theorem 4.3 illustrates the asymptotic expansion of $-\frac{d}{dt} \text{tr}(B^* \Gamma_t^{-1} B)$ in terms of the invariants \mathcal{I} and $\mathcal{Q}^{(i)}$. Its integral is

$$-\text{tr}(B^* \Gamma_t^{-1} B) = -\frac{\mathcal{N}}{t} + c + \sum_{i=0}^h \text{tr}(\mathcal{Q}^{(i)}) \frac{t^{i+1}}{i+1} + O(t^{h+2})$$

for some constant c , coming from the integration. Recall that $D_t = -\Gamma_{-t}$. Thus in Eq. (19) we have

$$\frac{f'(t)}{f(t)} = 2\text{tr}(A) + c + \sum_{i=0}^h (-1)^{i+1} \text{tr}(\mathcal{Q}^{(i)}) \frac{t^{i+1}}{i+1} + O(t^{h+2}).$$

Then we can write the determinant of D_t in the following exponential form depending on the invariants $\mathcal{Q}^{(i)}$

$$\det D_t = c_0 t^{\mathcal{N}} e^{(c+2\text{tr} A)t + \sum_{i=0}^h (-1)^{i+1} \text{tr} \mathcal{Q}^{(i)} \frac{t^{i+2}}{(i+1)(i+2)} + O(t^{h+3})},$$

for some constant c_0 and the constant $c := \text{tr}(C)$. In particular, one can easily find the first terms of the Taylor expansion of $\det D_t$ at $t = 0$, namely

$$(20) \quad \det D_t = c_0 t^{\mathcal{N}} \left[1 + (c + 2\text{tr} A)t + \frac{(c + 2\text{tr} A)^2}{2} t^2 - \frac{\text{tr} \mathcal{Q}^{(0)}}{2} t^2 + o(t^2) \right],$$

determined up to a constant c .

Remark 5.2. Notice that, in order to have a non degenerate matrix D_t , it is essential that the order \mathcal{N} of $\det D_t$ is the same term that appears in the asymptotics of $\text{tr} \mathcal{Q}(t)$.

5.1. The first term in the expansion. To find the constant c in Eq. (20) let us compute the first terms of the expansion of $\det D_t$. The derivative matrix $\dot{\Gamma}_t = e^{-tA} B B^* e^{-tA^*}$ is positive semi-definite and can be written as

$$\dot{\Gamma}_t = V(t) V(t)^*,$$

for $V(t) = e^{-tA}B$. Let $v_i(t)$ denote the columns of $V(t)$ and define the filtration $E_1 \subset E_2 \subset \dots \subset E_m = \mathbb{R}^n$ as

$$(21) \quad E_i = \text{span}\{v_j^{(l)}(0), 1 \leq j \leq k, 0 \leq l \leq i-1\}.$$

Therefore E_i is the subspace of \mathbb{R}^n defined by the columns of the matrices $A^j B$ for $0 \leq j \leq i-1$ and has dimension k_i . Choose coordinates on \mathbb{R}^n adapted to this filtration, i.e., associated with a basis $\{e_1, \dots, e_n\}$ of \mathbb{R}^n such that $\text{span}\{e_1, \dots, e_{k_i}\} = E_i$. In these coordinates $V(t)$ has a peculiar structure, namely

$$(22) \quad V(t) = \begin{pmatrix} \widehat{v}_1 \\ t\widehat{v}_2 \\ \vdots \\ t^{m-1}\widehat{v}_1 \end{pmatrix} + \begin{pmatrix} t\widehat{w}_1 \\ t^2\widehat{w}_2 \\ \vdots \\ t^m\widehat{w}_1 \end{pmatrix} + \begin{pmatrix} O(t^2) \\ O(t^3) \\ \vdots \\ O(t^{m+1}) \end{pmatrix},$$

where \widehat{v}_i and \widehat{w}_i are $(k_i - k_{i-1}) \times k$ constant matrices and every \widehat{v}_i has maximal rank. Let $\widehat{V}(t)$ and $\widehat{W}(t)$ be the first and second principal parts of $V(t)$, then the Taylor series of the matrix Γ_t can be found as

$$\Gamma_t = \int_0^t V(\tau)V(\tau)^* d\tau = \int_0^t \widehat{V}(\tau)\widehat{V}(\tau)^* + \left(\widehat{V}(\tau)\widehat{W}(\tau)^* + \widehat{W}(\tau)\widehat{V}(\tau)^*\right) d\tau + r(t),$$

where $r(t)$ is a remainder term. In components we write Γ_t as a $m \times m$ block matrix, whose block $\Gamma_{ij}(t)$ is the $(k_i - k_{i-1}) \times (k_j - k_{j-1})$ matrix with Taylor expansion

$$\begin{aligned} \Gamma_{ij}(t) &= \left(\frac{\widehat{v}_i \widehat{v}_j^*}{i+j-1}\right) t^{i+j-1} + \left(\frac{\widehat{v}_i \widehat{w}_j^* + \widehat{w}_i \widehat{v}_j^*}{i+j}\right) t^{i+j} + O(t^{i+j+1}) \\ &= \mathcal{X}_{ij} t^{i+j-1} + \mathcal{Y}_{ij} t^{i+j} + O(t^{i+j+1}), \end{aligned}$$

where \mathcal{X} and \mathcal{Y} are $m \times m$ block matrices implicitly defined by this formula. Moreover \mathcal{X} is invertible. Let $J_{\sqrt{t}}$ be the $n \times n$ diagonal matrix whose j -th element is equal to \sqrt{t}^{2i-1} for $k_{i-1} < j \leq k_i$, then

$$(23) \quad \Gamma_t = J_{\sqrt{t}} (\mathcal{X} + t\mathcal{Y} + O(t^2)) J_{\sqrt{t}}.$$

As expected the determinant of Γ_t is of order $t^{\mathcal{N}}$ and its Taylor expansion is computed in terms of \mathcal{X} and \mathcal{Y} as

$$\det \Gamma_t = t^{\mathcal{N}} \det(\mathcal{X})(1 + \text{tr}(\mathcal{X}^{-1}\mathcal{Y})t + o(t)).$$

Now we are ready to find the determinant of D_t . This follows from identity $D_t = e^{tA}\Gamma_t e^{tA^*} = -\Gamma_{-t}$. Indeed on the one hand the determinant of D_t is given by

$$(24) \quad \begin{aligned} \det D_t &= \det(e^{tA}\Gamma_t e^{tA^*}) = \det(e^{2tA}) \det \Gamma_t \\ &= \det(\mathcal{X}) t^{\mathcal{N}} [1 + (2\text{tr}A + \text{tr}(\mathcal{X}^{-1}\mathcal{Y}))t + o(t)]. \end{aligned}$$

On the other hand

$$(25) \quad \begin{aligned} \det D_t &= \det(-\Gamma_{-t}) = (-1)^n (-t)^{\mathcal{N}} \det(\mathcal{X}) [1 - t \text{tr}(\mathcal{X}^{-1}\mathcal{Y}) + o(t)] \\ &= \det(\mathcal{X}) t^{\mathcal{N}} [1 - \text{tr}(\mathcal{X}^{-1}\mathcal{Y})t + o(t)]. \end{aligned}$$

The last identity follows since n and \mathcal{N} have the same parity, indeed

$$\mathcal{N} = \sum_{i=1}^m (2i-1)k_i = 2M - n,$$

where $M = \sum_{i=1}^m ik_i$. Comparing equations (24) and (25) we find $\text{tr}(\mathcal{X}^{-1}\mathcal{Y}) = -\text{tr}A$, that means

$$\det D_t = \det(\mathcal{X}) t^{\mathcal{N}} (1 + (\text{tr}A)t + o(t)).$$

It follows that in formula (20) we have $c = -\text{tr}A$ and $c_0 = \det(\mathcal{X}) > 0$. This allows us to conclude that the asymptotics of the fundamental solution in $x = y \in \ker A$ for small time is

$$\begin{aligned}
 p(t, x, x) &= \frac{1}{(2\pi)^{n/2} \sqrt{\det D_t}} \\
 (26) \quad &= \frac{t^{-N/2}}{(2\pi)^{n/2} \sqrt{c_0}} e^{-\frac{\text{tr}A}{2}t} e^{\frac{1}{2} \sum_{i=0}^h (-1)^i \text{tr} Q^{(i)} \frac{t^{i+2}}{(i+1)(i+2)} + O(t^{h+3})} \\
 &= \frac{t^{-N/2}}{(2\pi)^{n/2} \sqrt{c_0}} \left(1 + \sum_{i=1}^h a_i t^i + O(t^{h+1}) \right),
 \end{aligned}$$

where the coefficients a_i can be explicitly computed from the expansion of the exponential. It follows that every a_i is polynomial in the components of $\text{tr}A$ and $\text{tr}Q^{(j)}$ for $j \leq i-2$ of order i and does not depend on x . In particular the first coefficients are computed as follows

$$\begin{aligned}
 a_1 &= -\frac{\text{tr}A}{2}, & a_2 &= \frac{(\text{tr}A)^2}{8} + \frac{\text{tr}Q^{(0)}}{4}, \\
 a_3 &= -\frac{\text{tr}Q^{(1)}}{12} - \frac{\text{tr}A \text{tr}Q^{(0)}}{8} - \frac{1}{48}(\text{tr}A)^3.
 \end{aligned}$$

5.2. Proof of Theorem 1.2. The proof of Theorem 1.2 is now an easy consequence of the analysis done until here.

Indeed by Remark 2.1, the geodesic that connects the point x to y in time t has starting associated covector $p_0 = \Gamma_t^{-1}(e^{-tA}y - x)$, therefore the minimal cost to go from x to y in time t is

$$S_t(x, y) = \frac{1}{2} (e^{-tA}y - x)^* \Gamma_t^{-1} (e^{-tA}y - x) = \frac{1}{2} (y - e^{tA}x)^* D_t^{-1} (y - e^{tA}x).$$

This is exactly the quantity appearing at the exponential of the heat kernel, that therefore can be written in terms of the optimal cost as

$$p(t, x, y) = \frac{e^{-S_t(x, y)}}{(2\pi)^{n/2} \sqrt{\det D_t}}.$$

The conclusion of the theorem is then a consequence of the Taylor expansion of $\frac{1}{\sqrt{\det D_t}}$, given in (26).

6. SMALL TIME ASYMPTOTICS OUT OF THE EQUILIBRIUM

In this section we prove Theorem 1.3, concerning the small time asymptotics of the fundamental solution at a point, x_0 , where the drift field is not zero.

By a translation of the origin, we can assume that $x_0 = 0$. This produces a no more linear, but affine drift field, i.e., $X_0(x) = Ax + \alpha$, where $\alpha := Ax_0$ is a column vector equal to the value of the drift at x_0 . Then we can study the original asymptotics at x_0 , through the asymptotics at the origin of the heat kernel of the linear pde, where to the drift field we add the constant value α . As shown in the Appendix its fundamental solution is

$$p(t, x, y) = \frac{e^{\varphi(t, x, y)}}{(2\pi)^{n/2} \sqrt{\det D_t}},$$

where

$$\varphi(t, x, y) = -\frac{1}{2} \left(y - e^{tA} \left(x + \int_0^t e^{-sA} ds \alpha \right) \right)^* D_t^{-1} \left(y - e^{tA} \left(x + \int_0^t e^{-sA} ds \alpha \right) \right)$$

and $D_t = e^{tA}\Gamma_t e^{tA^*}$ is the same covariance matrix as in (17). The original asymptotics of the fundamental solution at x_0 is given by the asymptotics of $(\det D_t)^{-1/2}$, found in the previous section, and the asymptotics of φ in $x = y = 0$, i.e.,

$$\varphi(t, 0, 0) = -\frac{1}{2}\alpha^* \left(\int_0^t e^{-sA} ds \right)^* \Gamma_t^{-1} \left(\int_0^t e^{-sA} ds \right) \alpha.$$

Let E_i be the subspaces defined in (21). If $\alpha = Ax_0 \in E_i \setminus E_{i-1}$, then $A^j \alpha \in E_{i+j}$ (actually it could possibly live in the previous subspaces, but not in a bigger one). Moreover $\int_0^t e^{-sA} ds = \sum_{i=1}^m -\frac{(-t)^i}{i!} A^{i-1} + O(t^{m+1})$. Therefore in coordinates adapted to the filtration $\{E_j\}_j$, the column vector $\int_0^t e^{-sA} ds \alpha$ can be written in m blocks of height d_j , as

$$\int_0^t e^{-sA} ds \alpha = \begin{pmatrix} t\alpha_1 \\ t\alpha_2 \\ \vdots \\ t\alpha_i \\ t^2\alpha_{i+1} \\ \vdots \\ t^{m-i+1}\alpha_m \end{pmatrix} + \begin{pmatrix} O(t^2) \\ O(t^2) \\ \vdots \\ O(t^2) \\ O(t^3) \\ \vdots \\ O(t^{m-i+2}) \end{pmatrix},$$

where α_j is a vector of length d_j for every $1 \leq j \leq m$ and α_i is not zero. The matrix Γ_t^{-1} can be written as a product of

$$\Gamma_t^{-1} = J_{1/\sqrt{t}}(\mathcal{X}^{-1} + O(t))J_{1/\sqrt{t}},$$

where $J_{1/\sqrt{t}}$ and \mathcal{X} are the matrices introduced in (23). Notice that the multiplication of $J_{1/\sqrt{t}}$ with $\int_0^t e^{-sA} ds \alpha$ gives the vector

$$J_{1/\sqrt{t}} \int_0^t e^{-sA} ds \alpha = \begin{pmatrix} \sqrt{t}\alpha_1 \\ \frac{1}{\sqrt{t}}\alpha_2 \\ \vdots \\ \frac{1}{\sqrt{t^{2i-3}}}\alpha_i \\ \frac{1}{\sqrt{t^{2i-3}}}\alpha_{i+1} \\ \vdots \\ \frac{1}{\sqrt{t^{2i-3}}}\alpha_m \end{pmatrix} + \begin{pmatrix} O(t\sqrt{t}) \\ O(\sqrt{t}) \\ \vdots \\ O(\frac{1}{\sqrt{t^{2i-1}}}) \\ O(\frac{1}{\sqrt{t^{2i-1}}}) \\ \vdots \\ O(\frac{1}{\sqrt{t^{2i-1}}}) \end{pmatrix}.$$

From here we see immediately that, since \mathcal{X}^{-1} has maximal rank and is positive definite, the scalar product

$$(J_{1/\sqrt{t}} \int_0^t e^{-sA} ds \alpha)^* \mathcal{X}^{-1} (J_{1/\sqrt{t}} \int_0^t e^{-sA} ds \alpha)$$

has a simple pole of order $-(2i-3)$ in 0. So if $\alpha \in E_i \setminus E_{i-1}$ with $i > 1$, then $\varphi(t, 0, 0)$ blows up for small t as $\frac{1}{t^{2i-3}}$, i.e., the heat kernel decreases with exponential velocity precisely as

$$p(t, x_0, x_0) = \frac{t^{-\mathcal{N}/2}}{(2\pi)^{n/2} \sqrt{c_0}} \exp\left(-\frac{C_1 + O(t)}{t^{2i-3}}\right) \quad \text{for } t \searrow 0, Ax_0 \in E_i \setminus E_{i-1},$$

for a positive constant C_1 . This proves the second part of Theorem 1.3.

Let us consider now the case $\alpha \in \text{span} B$. With this assumption the function $\varphi(t, 0, 0)$ is smooth in $t = 0$ and we want to find its exact value at the first order.

With a change of variables we can assume that the matrix B is the identity matrix in the first k rows and zero on the last $n - k$ rows. Moreover let $y \in \mathbb{R}^k$ be such that $Ax_0 = By$. We claim that

$$\varphi(t, 0, 0) = -t \frac{|y|^2}{2} + o(t) = -t \frac{|Ax_0|^2}{2} + o(t).$$

We can write the function φ as

$$\varphi(t, 0, 0) = -\frac{1}{2}y^* \left(\int_0^t e^{-sA} B ds \right)^* \Gamma_t^{-1} \left(\int_0^t e^{-sA} B ds \right) y$$

and we need to characterize $e^{-sA}B$. Using the results of the previous section, in coordinates adapted to the filtration we can write

$$\int_0^t e^{-sA} B ds = \begin{pmatrix} t\hat{v}_1 \\ t^2 \frac{\hat{v}_2}{2} \\ \vdots \\ t^m \frac{\hat{v}_m}{m} \end{pmatrix} + \begin{pmatrix} O(t^2) \\ O(t^3) \\ \vdots \\ O(t^{m+1}) \end{pmatrix},$$

where the \hat{v}_i are defined in (22). They are $d_i \times k$ matrix of maximal rank and \hat{v}_1 is the $k \times k$ identity matrix. We can highlight the dependence on t by multiplying on the left by the diagonal matrix K_t with j -th entry equal to t^i , for $k_{i-1} < j \leq k_i$. Also the matrix Γ_t^{-1} can be written as a product (see (23)), then

$$\int_0^t e^{-sA} B ds = K_t \left(\begin{pmatrix} \mathbb{I}_k \\ C \end{pmatrix} + O(t) \right) \quad \text{and} \quad \Gamma_t^{-1} = J_{1/\sqrt{t}} (\mathcal{X}^{-1} + O(t)) J_{1/\sqrt{t}},$$

where C is the $(n - k) \times k$ matrix composed by the last $n - k$ rows of the principal part of $\int_0^t e^{-sA} B ds$. Since $K_t \cdot J_{1/\sqrt{t}} = \sqrt{t} \mathbb{I}_n$, then

$$(27) \quad \varphi(t, 0, 0) = -\frac{t}{2} y^* \begin{pmatrix} \mathbb{I}_k \\ C \end{pmatrix}^* \mathcal{X}^{-1} \begin{pmatrix} \mathbb{I}_k \\ C \end{pmatrix} y + O(t^2).$$

Notice that $[\mathcal{X}]_{11} = \mathbb{I}_k$ and $[\mathcal{X}]_{i1} = \frac{\hat{v}_i}{i} = [\mathcal{X}]_{1i}^*$, for every $2 \leq i \leq m$, hence \mathcal{X} can be written as a block matrix

$$\mathcal{X} = \begin{pmatrix} \mathbb{I}_k & C^* \\ C & E \end{pmatrix},$$

where E is a $(n - k) \times (n - k)$ matrix.

Lemma 6.1. *The inverse matrix \mathcal{X}^{-1} is the block matrix*

$$\mathcal{X}^{-1} = \begin{pmatrix} [\mathcal{X}^{-1}]_{11} & [\mathcal{X}^{-1}]_{12} \\ [\mathcal{X}^{-1}]_{12}^* & [\mathcal{X}^{-1}]_{22} \end{pmatrix},$$

where

$$[\mathcal{X}^{-1}]_{11} = \mathbb{I}_k + C^* (E - CC^*)^{-1} C, \\ [\mathcal{X}^{-1}]_{12} = -C^* (E - CC^*)^{-1}, \quad [\mathcal{X}^{-1}]_{22} = (E - CC^*)^{-1}.$$

Proof. This is the general expression of the inverse of a block matrix, provided $[\mathcal{X}]_{11}$ and $E - CC^*$ are not singular. $[\mathcal{X}]_{11}$ is the identity matrix. Let us show that $E - CC^*$ is not singular. Assume $x \in \mathbb{R}^{n-k}$ satisfies $(E - CC^*)x = 0$. Then the column vector (of dimension n) equal to $-(C^*x)^*, x^*$ is in the kernel of \mathcal{X} . Therefore $x = 0$, since \mathcal{X} is not singular. \square

Applying the Lemma to Eq. (27) we find that

$$\begin{aligned} \varphi(t, 0, 0) &= -\frac{t}{2} y^* ([\mathcal{X}^{-1}]_{11} + ([\mathcal{X}^{-1}]_{12} C)^* + [\mathcal{X}^{-1}]_{12} C + C^* [\mathcal{X}^{-1}]_{22} C) y + O(t^2) \\ &= -\frac{t}{2} y^* \mathbb{I}_k y + O(t^2) = -\frac{|Ax_0|^2}{2} t + O(t^2). \end{aligned}$$

Taking into account the asymptotics of $(\det D_t)^{-1/2}$ found in the previous section, this completes the proof of Theorem 1.3.

APPENDIX A. ON THE FUNDAMENTAL SOLUTION

In the following section we derive the fundamental solution of the linear partial differential equation

$$(28) \quad \frac{\partial \varphi}{\partial t} - \sum_{j=1}^n (\alpha + Ax)_j \frac{\partial \varphi}{\partial x_j} - \frac{1}{2} \sum_{j,h=1}^n (BB^*)_{jh} \frac{\partial^2 \varphi}{\partial x_j \partial x_h} = 0, \quad \text{for } \varphi \in C_0^\infty(\mathbb{R} \times \mathbb{R}^n),$$

where α is a constant column vector of dimension n , A is a $n \times n$ real matrix, which represents the linear part of a *drift* field, and B is a $n \times k$ real matrix, with $1 \leq k \leq n$, that generates the diffusion coefficient.

An operator $\partial_t - L$ is hypoelliptic if for every open set $U \in \mathbb{R}^n$ and for every function φ such that $(\partial_t - L)\varphi \in C^\infty(U)$, then also $\varphi \in C^\infty(U)$. Hörmander proved in [13] that if Kalman's condition on the matrices A, B holds, that is

$$(29) \quad \text{rk}[B, AB, A^2B, \dots, A^{n-1}B] = n,$$

then the operator in (28) is hypoelliptic. Moreover in this case the operator admits a well defined fundamental solution, $p(t, x, y)$, that is given by the probability density of the solution ξ_t of the associated stochastic differential equation

$$(30) \quad \begin{cases} d\xi_t = (\alpha + A\xi_t)dt + Bdw(t) \\ \xi_0 = x \end{cases}$$

where $w(t) = (w_1(t), \dots, w_k(t))$ is a k -dimensional Brownian motion. In other words, $p(t, x, y)$ is the $C^\infty(\mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n)$ function such that for every Borel set $U \in \mathbb{R}^n$ the probability of ξ_t to be in U at time t is given by

$$P[\xi_t \in U | \xi_0 = x] = \int_U p(t, x, y) dy.$$

Here \mathbb{R}^+ denotes the strictly positive real line. The stochastic process solution of (30) is equal to

$$\xi_t = e^{tA} \xi_0 + e^{tA} \int_0^t e^{-sA} ds \alpha + e^{tA} \int_0^t e^{-sA} B dw(s),$$

as it can be readily verified with a derivation. In particular, if the initial value ξ_0 is Gaussian distributed, then ξ_t is a Gaussian process, since by definition of solution ξ_0 and $w(t)$ are independent. The initial condition is a Dirac delta centered in x , so it is a degenerate Gaussian with mean x and vanishing covariance matrix, therefore to find the distribution of ξ_t it is enough to determine its mean value and covariance matrix. To this end let us recall Itô's differential formula: let $\psi(t, x)$ be a smooth function, then for every $0 \leq s < t$

$$\begin{aligned} \psi(t, \xi_t) - \psi(s, \xi_s) &= \int_s^t \frac{\partial \psi(\tau, \xi_\tau)}{\partial t} d\tau \\ &+ \int_s^t \sum_{j=1}^n (\alpha + A\xi_\tau)_j \frac{\partial \psi(\tau, \xi_\tau)}{\partial x_j} + \frac{1}{2} \sum_{j,h=1}^n (BB^*)_{jh} \frac{\partial^2 \psi(\tau, \xi_\tau)}{\partial x_j \partial x_h} d\tau \\ &+ \int_s^t \sum_{i=1}^k \sum_{j=1}^n B_{ji} \frac{\partial \psi(\tau, \xi_\tau)}{\partial x_j} dw_i(\tau). \end{aligned}$$

To find the mean value of ξ_t let us fix the auxiliary function $\psi := x_j$ for $1 \leq j \leq n$, then

$$\xi_j(t) - \xi_j(0) = \int_0^t (\alpha + A\xi(\tau))_j d\tau + \int_0^t \sum_{i=1}^k B_{ji} dw_i(\tau).$$

Let $m_j(t)$ be the j -th component of the mean value of ξ_t . By the previous formula, since the integral of the Brownian motion has zero mean value, we find that $m(t)$ satisfies the differential equation $\dot{m}_j(t) = (\alpha + Am(t))_j$ with initial condition $m(0) = x$. Therefore

$$E[\xi_t | \xi_0] = e^{tA} \left(x + \int_0^t e^{-sA} ds \alpha \right).$$

To compute the covariance matrix let us choose $\psi = (x_j - m_j(t))(x_h - m_h(t))$, then by Itô differential formula it holds

$$\begin{aligned} \psi(t, \xi(t)) - \psi(0, \xi(0)) &= \int_0^t -\dot{m}_j(\tau)(\xi_h(\tau) - m_h(\tau)) - \dot{m}_h(\tau)(\xi_j(\tau) - m_j(\tau)) d\tau \\ &\quad + \int_0^t (\alpha + A\xi)_j(\xi_h(\tau) - m_h(\tau)) + (\alpha + A\xi)_h(\xi_j(\tau) - m_j(\tau)) + (BB^*)_{jh} d\tau \\ &\quad + \int_0^t \sum_{i=1}^k B_{ji}(\xi_h(\tau) - m_h(\tau)) + B_{hi}(\xi_j(\tau) - m_j(\tau)) dw_i(\tau). \end{aligned}$$

Let us take the expectation value and denote by $\rho_{jh}(t)$ the jh -component of the covariance matrix. Recall that $E[(\xi_j(t) - m_j(t)) | \xi(0)] = 0$. To evaluate

$$E[(A\xi(\tau))_j(\xi_h - m_h(\tau)) | \xi(0)]$$

we rewrite it as

$$E[(A(\xi(\tau) - m(\tau)))_j(\xi_h - m_h(\tau)) + (Am(\tau))_j(\xi_h - m_h(\tau)) | \xi(0)] = (A\rho(\tau))_{jh} + 0.$$

Then $\rho(t)$ satisfies the differential equation $\dot{\rho}_{jh}(t) = (A\rho(t))_{jh} + (A\rho(t))_{hj} + (BB^*)_{jh}$ with vanishing initial value, whose solution is

$$D_t := E[(\xi_j - m_j(t))(\xi_h - m_h(t)) | \xi(0)] = e^{tA} \int_0^t e^{-\tau A} BB^* e^{-\tau A^*} d\tau e^{tA^*}.$$

By Kalman's condition (29), the matrix D_t is invertible for every $t > 0$. Therefore we can conclude that the C^∞ fundamental solution of equation (28) is given by the non-degenerate Gaussian

$$p(t, x, y) = \frac{e^{\varphi(t, x, y)}}{(2\pi)^{n/2} \sqrt{\det D_t}},$$

where

$$\varphi(t, x, y) = -\frac{1}{2} \left(y - e^{tA} \left(x + \int_0^t e^{-sA} ds \alpha \right) \right)^* D_t^{-1} \left(y - e^{tA} \left(x + \int_0^t e^{-sA} ds \alpha \right) \right).$$

In the case when $\alpha = 0$ the formula for $p(t, x, y)$ reduces to

$$p(t, x, y) = \frac{e^{-\frac{1}{2}(y - e^{tA}x)^* D_t^{-1}(y - e^{tA}x)}}{(2\pi)^{n/2} \sqrt{\det D_t}} \quad \text{for } t > 0, x, y \in \mathbb{R}^n.$$

Acknowledgments. This research has been supported by the European Research Council, ERC StG 2009 “GeCoMethods”, contract number 239748 and by the ANR project SRGI “Sub-Riemannian Geometry and Interactions”.

REFERENCES

- [1] A. A. Agrachev, D. Barilari, and L. Rizzi. The curvature: a variational approach. *To appear on Memoires AMS*.
- [2] A. A. Agrachev and Y. Sachkov. *Control Theory from the Geometric Viewpoint*, volume 87. Springer-Verlag Berlin Heidelberg, 2004.
- [3] G. Ben Arous. Développement asymptotique du noyau de la chaleur hypoelliptique sur la diagonale. *Ann. Inst. Fourier (Grenoble)*, 39(1):73–99, 1989.
- [4] D. Barilari. Trace heat kernel asymptotics in 3d contact sub-riemannian geometry. *Journal of Mathematical Sciences*, 195(3):391–411, 2013.
- [5] D. Barilari, U. Boscain, G. Charlot, and R. W. Neel. On the heat diffusion for generic Riemannian and sub-Riemannian structures. *ArXiv e-prints*, October 2013.
- [6] D. Barilari, U. Boscain, and R. W. Neel. Small-time heat kernel asymptotics at the sub-Riemannian cut locus. *J. Differential Geom.*, 92(3):373–416, 2012.
- [7] F. Baudoin and J. Wang. The subelliptic heat kernel on the CR sphere. *Math. Z.*, 275(1-2):135–150, 2013.
- [8] F. Baudoin and J. Wang. The subelliptic heat kernels of the quaternionic Hopf fibration. *Potential Anal.*, 41(3):959–982, 2014.
- [9] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale (i). *Probability Theory and Related Fields*, 90:175–202, 1991.
- [10] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale (ii). *Probability Theory and Related Fields*, 90:377–402, 1991.
- [11] M. Berger, P. Gauduchon, and E. Mazet. *Le spectre d’une variété riemannienne*. Lecture Notes in Mathematics, Vol. 194. Springer-Verlag, Berlin, 1971.
- [12] J.-M. Bismut. *Large deviations and the Malliavin calculus*, volume 45 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1984.
- [13] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.
- [14] E. Lanconelli and S. Polidoro. On a class of hypoelliptic evolution operators. *Rend. Sem. Mat. Univ. Politec. Torino*, 52(1):29–63, 1994. Partial differential equations, II (Turin, 1993).
- [15] R. Léandre. Majoration en temps petit de la densité d’une diffusion dégénérée. *Probab. Theory Related Fields*, 74(2):289–294, 1987.
- [16] R. Léandre. Minoration en temps petit de la densité d’une diffusion dégénérée. *J. Funct. Anal.*, 74(2):399–414, 1987.
- [17] S. A. Molčanov. Diffusion processes and Riemannian geometry. *Uspehi Mat. Nauk*, 30(1(181)):3–59, 1975.
- [18] R. Neel and D. Stroock. Analysis of the cut locus via the heat kernel. In *Surveys in differential geometry. Vol. IX*, Surv. Differ. Geom., IX, pages 337–349. Int. Press, Somerville, MA, 2004.
- [19] E. Paoli. Small time asymptotics on the diagonal for Hörmander’s type hypoelliptic operators. *ArXiv e-prints*, February 2015.
- [20] S. Rosenberg. *The Laplacian on a Riemannian Manifold*, volume 31 of *London Mathematical Society Student Texts*. Cambridge University Press, 1997.
- [21] N. K. Stanton and D. S. Tartakoff. The heat equation for the $\bar{\partial}_b$ -Laplacian. *Comm. Partial Differential Equations*, 9(7):597–686, 1984.
- [22] D. W. Stroock and S. R. S. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 3:333–359, 1972.
- [23] S. R. S. Varadhan. On the behavior of the fundamental solution of the heat equation with variable coefficients. *Comm. Pure Appl. Math.*, 20:431–455, 1967.